Opinion TRENDS in Microbiology Vol.13 No.11 November 2005

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

"Reprinted with permission from Elsevier"

Papillomaviruses: different genes have different histories

Santiago García-Vallvé[1], Ángel Alonso[2] and Ignacio G. Bravo[3]

[1]Evolutionary Genomics Group, Biochemistry and Biotechnology Department, Rovira i Virgili University (URV), c/ Marcel-li Domingo, s/n. Campus Sescelades, 43007 Tarragona, Spain
[2]Deutsches Krebsforschungszentrum, F050, Im Neuenheimer Feld-242, 69120 Heidelberg, Germany
[3]Deutsches Krebsforschungszentrum, F050, Im Neuenheimer Feld-242, 69120 Heidelberg, Germany

Papillomaviruses (PVs) infect stratified squamous epithelia in vertebrates. Some PVs are associated with different types of cancer and with certain benign lesions. It has been assumed that PVs coevolved with their hosts. However, recently it has been shown that different regions of the genome have different evolutionary histories. The PV genome has a modular nature and appeared after the addition of pre-existent blocks. This order of appearance in the PV genome is evident today in the different evolutionary rates of the different genes, with new genes – E5, E6 and E7 – diverging faster than old genes – E1, E2, L2 and L1. Here, we propose an evolutionary framework aiming to integrate genome evolution, PV biology and epidemiology of PV infections.

Classification of PVs

Papillomaviridae are a family of small dsDNA viruses that infect warm-blooded vertebrates [1]. Members of this family are associated with all clinical cases of cervical cancer and are also found in other benign and malignant proliferative disorders, such as skin warts, genital warts, laryngeal papillomas and nonmelanoma skin cancer [2–5]. In all cases, the target cells of the papillomaviruses (PVs) are located in the basal layer of stratified squamous epithelia, cutaneous or mucosal [6].

PVs have been classified according to the sequence identity in one of the capsid proteins, L1, without taking into account any other region of the genome and without considering phenotypic characteristics of the viral infection [7]. This classification based on L1 has led to the proliferation of discrete PV genera that comprise single species, yielding a taxonomy that provides only descriptive information [7]. Despite the growing knowledge about the infection and tumorigenesis processes, no comprehensive model bringing together the biology and manifestations of the infection of PVs has yet been provided. Recent studies have shown the first links between the chemistry of the viral proteins and the epidemiology of the infection in alpha PVs, based on viral evolution [8–10]. The alpha PVs mainly cause mucosal lesions in primates, although some of them are also associated with cutaneous warts [7]. Some alpha PVs, such as HPV16 or HPV18, are causative agents of cervical cancer and are referred to as high-risk viruses. Other alpha PVs, such as HPV6 or HPV11, give rise to non-malignant proliferative disorders and are therefore referred to as low-risk viruses [3,11]. The phylogenies of capsid proteins and of transforming proteins are different in alpha PVs [8,9]. Thus, high-risk alpha PVs appear together regarding the phylogeny of the transforming proteins but do not appear together according to the phylogeny of the capsid proteins [9]. Additionally, the number of accepted mutations per position is higher in the oncogenes than in the rest of the PV genes [9] and there is evidence of positive Darwinian selection in the oncogenes but not in the rest of the PV genes [8].

The aim of the present work is to provide a comprehensive framework for the evolutionary history of the PVs. On the one hand, the evolutionary hypotheses must explain the growing knowledge of the ubiquity and diversity of the PVs. On the other hand, the phylogenetic inferences must justify the variations in anatomical tropism and association to tumorigenesis in different PVs.

The basic structure of PVs

The basic structure of the PV genome is depicted in Figure 1. The only elements shared by all the members of the *Papillomaviridae* family are the presence of an upstream regulatory region (URR), the early proteins E1 and E2, and the late proteins L1 and L2. The URR is not transcribed and contains enhancer and promoter sequences. The E1 protein is a helicase involved in viral episomal replication [12]. The E2 protein is a transcription factor with a self-inhibiting dual activity. In the initial phases of infection, E2 is expressed at low levels and enhances the viral transcription allowing the expression of the E1 and E2 proteins. As the infected cell differentiates, E2 is expressed at higher levels and represses transcription from the early promoters, thus arresting the expression of the early proteins [13,14]. The L1 protein is the main component of the viral icosahedrical capsid and is capable of auto-assembly, thus giving rise to viral-like particles [15]. Finally, the L2 protein binds the circular viral DNA and packages it into the viral capsid [16]. These four proteins alone are theoretically able to fulfil the basic tasks of replicating, regulating, stabilizing and packaging the viral DNA, which lead ultimately to the release of the virion progeny [17].

Corresponding author: Bravo, I.G. (i.bravo@dkfz.de).
Available online 21 September 2005

 0966-842X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tim.2005.09.003

Apart from the highly conserved structure described previously, there are two regions in the PV genome that display a high variability and show a broad repertoire of encoded open reading frames (ORFs): the E6–E7 region, located between the URR and the E1 ORF, and the hinge region between the E2 and L2 ORFs.

The E6–E7 region of the PV genome

The E6–E7 region is present in all PVs and contains genes that encode transforming proteins that are expressed in the initial stages of the infection, before the infected cell starts to differentiate [1,17]. These proteins have been given the same names, according to their position in the genome (i.e. E6 and E7). However, this practice can lead to confusion because proteins with different sequences and properties bear the same name. There is little, if any, sequence homology between the E7 proteins in distant PVs. In addition, when measured in closely related viruses, the branch distances from the present E6 and E7 proteins to their ancestors are approximately two times higher than that of the structural proteins L1 and L2 [9] (Figures 1 and 2). High evolutionary distances lead to poor alignments and phylogenetic reconstruction is based on sequence alignments. Therefore, the phylogenetic support (i.e. the confidence levels of the different nodes of the phylogenetic trees) is higher for the late proteins than for the early proteins [9] (S. Garcia-Vallve *et al.*, unpublished). Furthermore, in mammalian genomes, older genes have evolved more slowly than more recent genes [18]. If this were also true for the PVs, the E6 and E7 genes could have gained access to the genome after the core region was formed. This suggests that PVs originated via a multistep process that generated the present genome organization through the addition of blocks previously present in other organisms.

The E7 proteins from HPV16, CRPV or BPV1 have transforming properties [1,19,20]. They bind pRb and prevent its interaction with the transcription factor E2F-1 [21]. This leads to expression of E2F-1 responsive genes, thus resulting in cell growth. Other E7 proteins (e.g. HPV10 E7, HPV20 E7 or RPV E7), however, do not interact *in vitro* with pRb. Therefore, E7 proteins from PVs that infect closely related hosts do not necessarily show the same biological activities [22–24]. The same picture is true for the E6 proteins. The E6 proteins from human PVs involved in cervical cancer target the tumour suppressor protein p53 for degradation, via ubiquitination [25,26]. By contrast, BPV1 E6 protein also shows transforming activity but does not bind to p53 [26,27]. Furthermore, the E6 ORF is not present in all PVs: avian PVs [28,29] and porpoise PVs display a long ORF in the locus where E6 and E7 genes reside in alpha or beta PVs (Figure 1).

Different PVs infecting the same host also display different genome arrangements within the E6–E7 region. As an example, BPV1 and BPV5 present the typical arrangement with E6–E7 in tandem, whereas BPV3 and BPV4 lack the E6 gene and instead have a short hydrophobic ORF. This ORF was originally termed E8 but was later renamed E5 because it shows the characteristics of the E5 ORFs of delta PVs [30]. The presence of an ORF that encodes a short hydrophobic peptide in the E6–E7 region is an additional feature of some PV genomes, which might have gone unnoticed in previous analyses. Such short hydrophobic ORFs appear either to replace the E6 gene in certain bovine PVs, to overlap the E6 ORF in other PVs infecting deer, humans and rabbits, or to overlap the E7 gene in PVs infecting birds (Figure 1).

The E2–L2 region of the PV genome

The E2–L2 region of the PV genome codes for the different members of the E5 family in alpha PVs and in delta PVs [31] (Figure 1). The E5 proteins in alpha PVs can be classified into four different groups according to their hydrophobic profiles and phylogeny. The presence of a given E5 type correlates with the clinical manifestations of the corresponding viral infection: viruses that contain E5α are associated with cervical cancer; viruses that contain E5γ and/or E5δ are associated with mucosal benign lesions and viruses that contain E5β are associated with benign cutaneous lesions [9]. The expression of the E5 protein of human PVs associated with cervical cancer alters different membrane-related processes, such as major histocompatibility complex trafficking to the cell surface, endosomal acidification or induction of ligand-mediated apoptosis [31]. The initial perturbation caused by E5 could possibly occur through changes in the composition and dynamics of the cell membranes [32]. The E5 protein from BPV1 is a true oncogene, which stimulates ligand-independent cell growth [27]. Despite sharing the same name, the E5 proteins from alpha and delta PVs share only an overall hydrophobic nature but no sequence homology [9].

Many PVs do not contain an E2–L2 region, as is the case for cutaneous human PVs from genera beta, gamma and mu, for PVs infecting rabbits and for all orphan PVs. Lambda PVs seem to be a special case. With regards to other regions of the genome lambda PVs are closely related to kappa and mu genera [33]. However, lambda PVs present a long E2–L2 region, which spans 1 kb without containing any coding sequence [34]. The functions of this region are unknown, and its presence and conservation pose an interesting evolutionary mystery.

Different evolutionary histories

Apart from the differences in activities between proteins with the same name and in genome arrangement, the phylogenetic relationships across different regions of the PV genome are not the same (Figure 2). The members of certain genera always cluster together confidently, independently of which region of the genome is analyzed. This is the case for species belonging to genera alpha, beta, gamma, delta, kappa, lambda and mu.

Genera beta and gamma infect skin in primates and are consistently closely related to each other. BPV4 also clusters together with beta and gamma PVs according to E1, E2, L2 and L1 phylogenies but not according to E6 and E7 phylogenies because BPV4 lacks an E6 ORF. Thus, BPV4 and PVs beta and gamma could have shared a common ancestor for the core region E1–E2–L2–L1 and could be regarded as a supergenus. Genera kappa, lambda

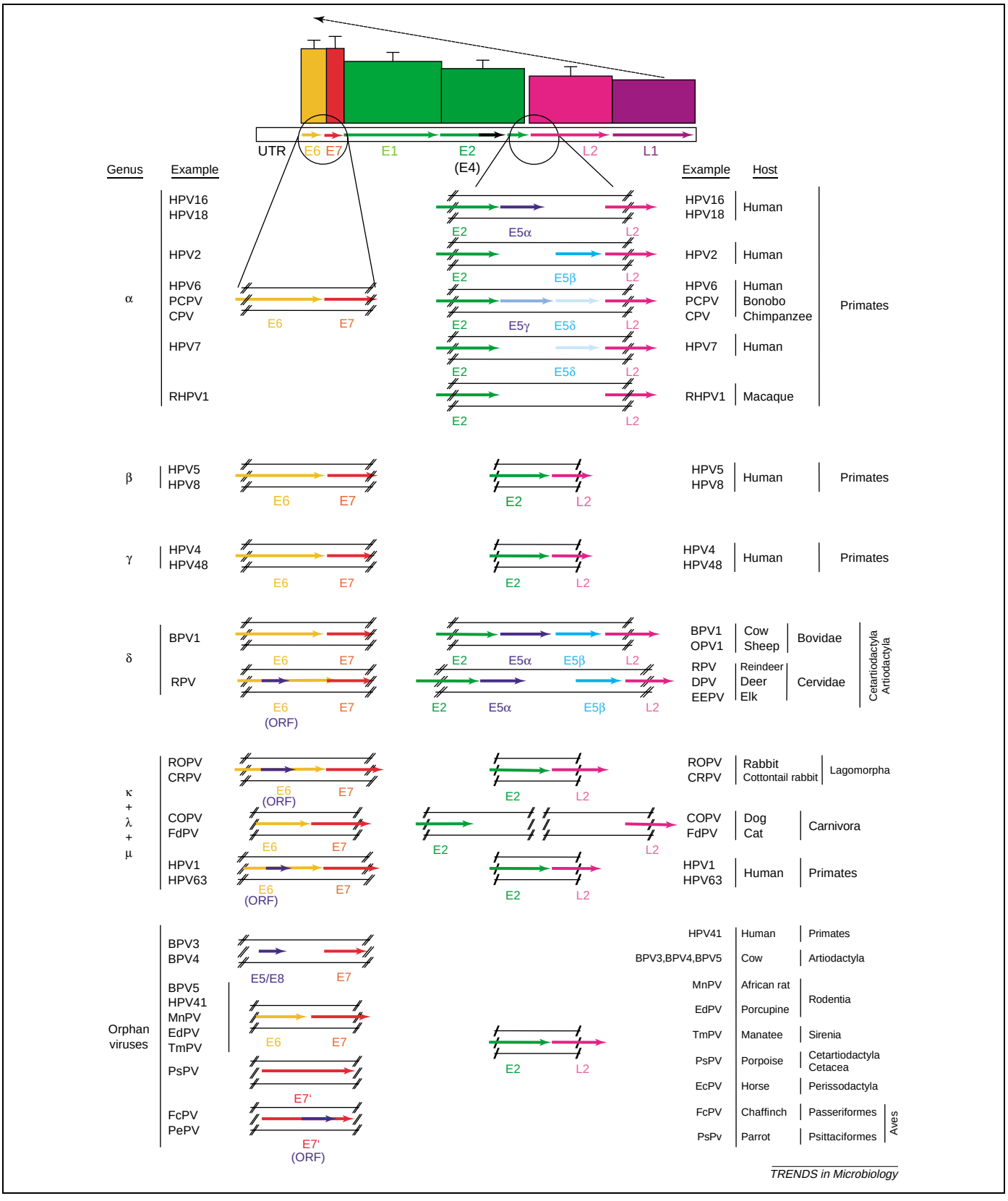


Figure 1. Genome arrangement and relative divergence in the *Papillomaviridae* family. Papillomaviruses (PV) are classified into genera named following the Greek alphabet attending to the homology in the L1 ORF (left columns). PV types are named after the host name, either in English or in Latin, as follows: BPV, bovine PV; OPV, ovine PV; RPV, reindeer PV; DPV, deer PV; EEPV, European elk PV; PsPV, *Phocoena spinipinnis* (Burmeister's porpoise) PV; COPV, canine oral PV; FdPV, *Felis domesticus* (cat) PV; EcPV, *Equus caballus* (horse) PV; TmPV, *Trichechus manatus latirostris* (manatee) PV; ROPV, rabbit oral PV; CRPV, cottontail rabbit PV; HPV, human PV; PCPV, pigmy chimpanzee PV; CPV, chimpanzee PV; RHPV, rhesus monkey PV;MnPV, *Mastomys natalensis* (African-soft furred rat) PV; EdPV, *Erethizon dorsatum* (porcupine) PV; FcPV, *Fringilla coelebs* (chaffinch) PV; PsPV, *Psittacus erithacus* (grey parrot) PV. Common characters to all PV members are the URR region, containing enhancer and promoter sequences, the E1–E2 region, coding for proteins related to viral replication and the L2–L1 region, coding for capside proteins. Between URR and E1–E2 is the E6–E7 segment. This region displays different arrangements in different viruses and encodes for E6, E7, E7′ and E8–E5 ORFs, with transforming capacities. The hinge region between E2 and L2 is not present in all PVs and encodes for the different members of the E5 family, also with transforming ability. PVs infecting the same or closely related hosts do not necessarily

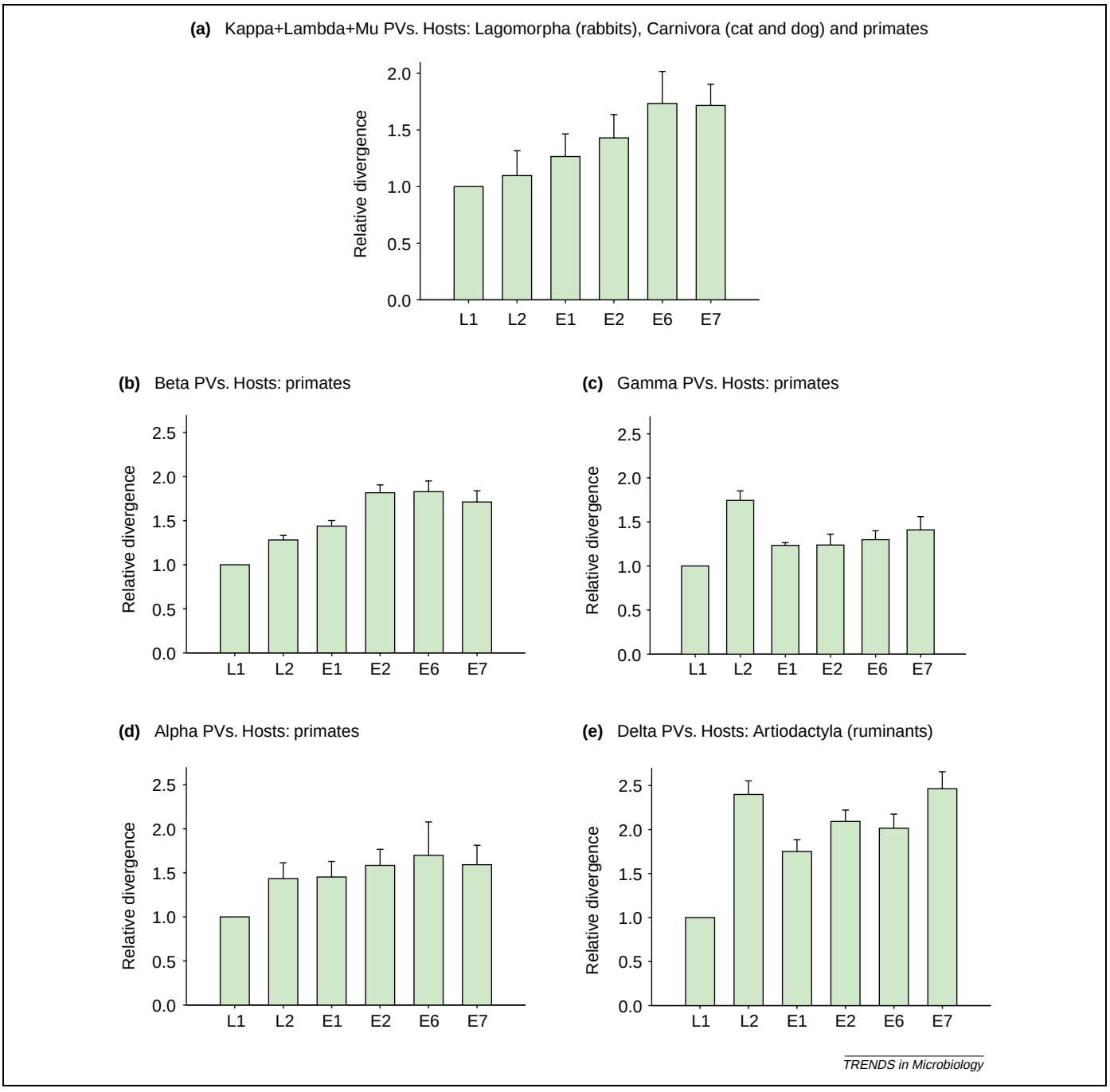
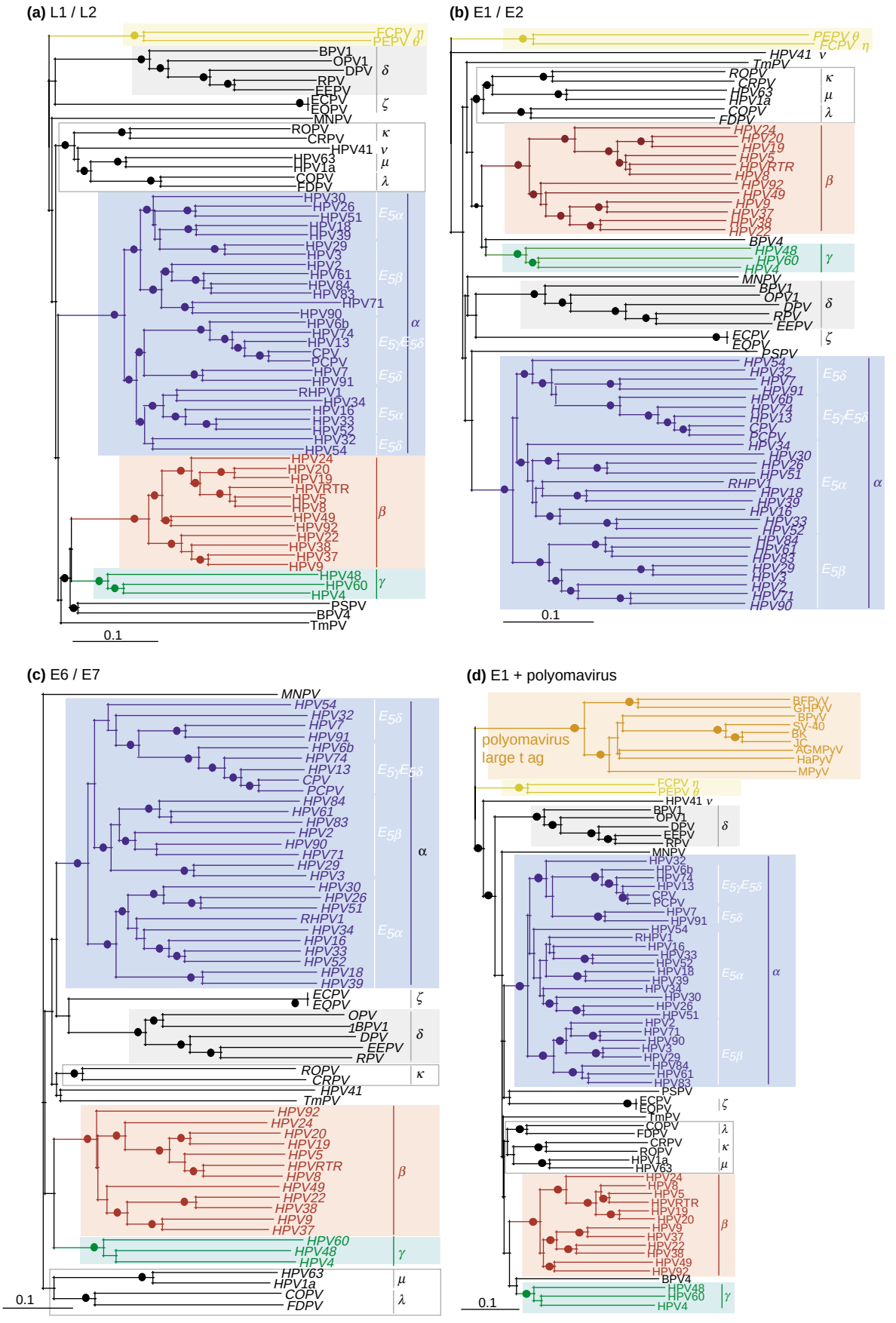


Figure 2. Early proteins have diverged more than late proteins in Papillomaviruses. Branch lengths in the trees of PV proteins from kappa+lambda+mu, beta, gamma, alpha and delta genus. Relative distances were measured from present PV proteins to the corresponding ancestral node of each genus were calculated and normalized with respect to L1 divergence. There is an obvious gradient of relative divergence, in the order L1-L2-E1-E2-E6-E7 in virtually all PVs. If the ancestral PVs of each genus is assumed to contain the same genes as current PVs, E6 and E7 proteins (the latest to arrive to the PVs genome in evolutionary terms) show the highest divergences rates. Note also the abnormal high divergence of L2 in gamma and delta PVs. Genes that have accessed recently to a genome show usually higher evolutionary rates than older genes [18]. This fact could reflect the evolutionary pressures considering sequence itself, shifting for instance from the original codon usage in the genome of origin to the codon usage in the recipient genome [49]. The different divergence of each PV protein reflects different functional constraints, probably related to their role in the virus infectious cycle. Relative variations in the divergence rates are also informative of putative additional pressures for a given ORF [35,50]. Thus, the high divergences of the L2 ORF in gamma PVs and in delta PVs could reflect additional functions or additional immune pressures in the infections associated to these viruses, compared with sister viruses that infect the same hosts [9].

and mu also appear together in the phylogeny of the core region and could have shared a common ancestor, thus constituting another coherent supergenus of distantly related PVs that infect different mammalian hosts.

Delta PVs always branch together in the phylogenetic trees of each of the analyzed regions of the genome. Within delta PVs, viruses infecting Cervidae form a stable group in all regions analyzed (Figure 3), whereas PVs infecting

display the same genome arrangement and/or encode the same ORFs. A brief description of the host phylogeny is provided (right columns). Proteins encoded in the PV genome do not shown the same relative divergences. As exemplified for the genera kappa, lambda and mu (bar diagram, upper part), the capside proteins L1 and L2 have diverged less than the replicative proteins E1 and E2, whereas the transforming proteins E6 and E7 show the highest divergence rate. Together, these facts point towards a composite nature of the PV genome, with the core URR-E1–E2-L2–L1 having appeared first, and the transforming proteins having accessed later to the genome.


(a) L1 / L2
FCPV η
PEPV θ
BPV1
OPV1
DPV
RPV
EEPV
ECPV
EQPV
MNPV
ROPV
CRPV
HPV41
HPV63
HPV1a
COPV
FDPV
HPV30
HPV26
HPV51
HPV18
HPV39
HPV29
HPV3
HPV2
HPV61
HPV84
HPV83
HPV71
HPV90
HPV6b
HPV74
HPV13
CPV
PCPV
HPV7
HPV91
RHPV1
HPV34
HPV16
HPV33
HPV52
HPV32
HPV54
HPV24
HPV20
HPV19
HPVRTR
HPV5
HPV8
HPV49
HPV92
HPV22
HPV38
HPV37
HPV9
HPV48
HPV60
HPV4
PSPV
BPV4
TmPV
0.1
(b) E1 / E2
PEPV θ
FCPV η
HPV41 ν
TmPV
ROPV
CRPV
HPV63
HPV1a
COPV
FDPV
HPV24
HPV20
HPV19
HPV5
HPVRTR
HPV8
HPV92
HPV49
HPV9
HPV37
HPV38
HPV22
BPV4
HPV48
HPV60
HPV4
MNPV
BPV1
OPV1
DPV
RPV
EEPV
ECPV
EQPV
PSPV
HPV54
HPV32
HPV7
HPV91
HPV6b
HPV74
HPV13
CPV
PCPV
HPV34
HPV30
HPV26
HPV51
RHPV1
HPV18
HPV39
HPV16
HPV33
HPV52
HPV84
HPV61
HPV83
HPV29
HPV3
HPV2
HPV71
HPV90
0.1
(c) E6 / E7
MNPV
HPV54
HPV32
HPV7
HPV91
HPV6b
HPV74
HPV13
CPV
PCPV
HPV84
HPV61
HPV83
HPV2
HPV90
HPV71
HPV29
HPV3
HPV30
HPV26
HPV51
RHPV1
HPV34
HPV16
HPV33
HPV52
HPV18
HPV39
ECPV
EQPV
OPV
BPV1
DPV
EEPV
RPV
ROPV
CRPV
HPV41
TmPV
HPV92
HPV24
HPV20
HPV19
HPV5
HPVRTR
HPV8
HPV49
HPV22
HPV38
HPV9
HPV37
HPV60
HPV48
HPV4
HPV63
HPV1a
COPV
FDPV
0.1
(d) E1 + polyomavirus
polyomavirus
large t ag
BFPyV
GHPyV
BPyV
SV-40
BK
JC
AGMPyV
HaPyV
MPyV
FCPV η
PEPV θ
HPV41 ν
BPV1
OPV1
DPV
EEPV
RPV
MNPV
HPV32
HPV6b
HPV74
HPV13
CPV
PCPV
HPV7
HPV91
HPV54
RHPV1
HPV16
HPV33
HPV52
HPV18
HPV39
HPV34
HPV30
HPV26
HPV51
HPV2
HPV71
HPV90
HPV3
HPV29
HPV84
HPV61
HPV83
PSPV
ECPV
EQPV
TmPV
COPV
FDPV
CRPV
ROPV
HPV1a
HPV63
HPV24
HPV8
HPV5
HPVRTR
HPV19
HPV20
HPV9
HPV37
HPV22
HPV38
HPV49
HPV92
BPV4
HPV48
HPV60
HPV4
0.1
TRENDS in Microbiology

Bovidae cluster separately according to E6–E7 but not according to the genomic core region.

Members of the Alphapapillomavirus genus always appear together in all phylogenetic trees but the topologies of the trees vary depending on the region analyzed. The topology of the trees for the early transforming proteins – E5, E6 and E7 – and for the early replicative proteins – E1 and E2 – correspond to the epidemiology of the associated infections. There are three main branches, one encompassing high-risk viruses, one comprising low-risk viruses, and a third one enclosing alpha PVs causing cutaneous lesions [9,10] (Figure 3). However, the topology of the phylogenetic tree of the late proteins does not follow this pattern, and high-risk and low-risk viruses appear intertwined [9,35,36] (Figure 3). There is, therefore, a change in the phylogenetic relationships within alpha PVs along the genome. Such changes in the topology of phylogenetic trees as we move along an alignment are commonly used to detect recombinations [37,38]. The topologies of the trees in alpha PVs are incoherent [7,9,10,35,36] and this fact points towards the existence of an early recombination event within the L2–L1 region of the alpha PVs that could have lead to their particular phylogeny. Further evidence of the particular evolution of the L1–L2 genes in the alpha PVs is obvious in their nucleotide composition and codon usage. PVs have nucleotide composition and codon preferences that differ strongly from those of the host [39,40]. All the early genes from alpha PVs infecting humans share some similar compositional patterns but the L1 and L2 genes of alpha PVs have a different nucleotide composition, which does not correlate with changes in protein expression associated to differentiation in the host cell [39]. The importance of the differential association of certain PVs with cervical cancer cannot be overlooked and must be integrated into their classification criteria, as the International Committee of Taxonomy of Viruses (ICTV) suggests [41]. For this reason, and in case a recombination process had taken place, the use of divergence in the L1 proteins as the yardstick for PV classification might have been an unfortunate choice.

A history of *Papillomaviridae*

Until now, a coevolution of virus and host was tacitly assumed for PVs [42–45], although other authors have also suggested multiple origins for some PVs [33,36,46]. If the hypothesis of coevolution were correct, all PVs infecting a given host would appear together, sharing a recent common ancestor. The phylogenetic relationships between viruses could then be reconstructed stepwise following the phylogenetic relationships between their hosts. However, this is not the case. Present PVs infecting primates are not monophyletic [33] and have arisen from at least four different ancestors. These putative ancestral viruses would have given rise separately to genus alpha, to genera beta and gamma, to genera mu and to genera nu. The same scenario is depicted for PVs infecting Artiodactyla, with at least two ancestors, one for the delta PVs and another for the different orphan bovine PVs (BPVs) [33,36]. Furthermore, there has been a common ancestor for the supergenus kappa+lambda+mu, which infect Lagomorpha, Carnivores and Primates, respectively. However, the most recent common ancestor for these three mammalian orders lived ~92 million years ago and was also common to many other orders, such as Artiodactyla, Cetacea, Sirenia or Perissodactyla [47]. If there had been a virus–host coevolution, PVs infecting cow, porpoise, manatee and horse should also co-appear in the proposed supergenus. Finally, Carnivora and Perissodactyla shared a recent common ancestor, and also Cetacea and Artiodactyla shared a recent common ancestor, but this is not the case for the respective PVs that infect them.

The last common ancestor of mammals and aves lived ~250 million years ago [47]. The presence of PVs in mammals and aves therefore suggests that the appearance of PVs predated the split between both branches in the Amniota lineage. We propose that this proto-PV would have comprised the initial versions of the core URR-E1–E2-L2–L1. The respective original sequences probably already existed as separated entities in other ancestral organisms, as shown by the homology between E1 and the large T antigen in polyomaviruses [48]. The ancestral versions of the E6 and E7 genes present today in PVs infecting mammals would have gained access to the genome later, providing transforming capacities to the receiving genomes. The acquisition of the E5 genes in the E2–L2 region, and also the hydrophobic E5–E8 genes in the E6–E7 region might have occurred later. In the case of the alpha PVs the accession of the proto-E5 ORFs predated the split between high-risk and low-risk viruses, because all high-risk viruses bear the same type of E5, namely E5α, whereas all low-risk viruses code for either E5β, E5γ or E5δ [9]. The last common ancestor to both high-risk and low-risk viruses predated the divergence of the ancestors of the *Maccaca* and *Homo* genera because rhesus monkey PV clusters together with human high-risk PVs, and chimpanzee and bonobo PVs cluster together with human low-risk PVs. However, the non-primate counterparts of the alpha PVs, causing mucosal lesions in their hosts, are still to be found and no information about the evolution of the oncogenes before the appearance of primates can be inferred. Therefore, a thorough study of PVs in genital and cutaneous lesions in

Figure 3. Phylogenetic trees of Papillomaviruses (PV). The trees were constructed from computationally fused proteins **(a)** L1 and L2, **(b)** E1 and E2 and **(c)** E6 and E7. From these trees, the hypothesis of a sustained coevolution between PVs and their hosts is not supported. L1–L2 and E1–E2 trees were rooted between the node that separates avian (FCPV and PEPV) and mammalian PVs. This position was chosen according to the taxonomic relationships of hosts and it was confirmed by **(d)** a phylogenetic tree of the E1 proteins using a seed of large T antigen sequences from polyomaviruses as an outgroup. Protein sequences were aligned with clustalw and phylogenetic trees were constructed using the neighbour-joining algorithm. Topologically similar trees were obtained using a maximum likelihood method and different alignment algorithms. Circles at nodes indicate bootstrap support values above 600 in 1000 replicates. Well-defined clusters, independent of the region analyzed, like genera alpha, beta, gamma, delta and kappa, lambda and mu are shaded or boxed. The four groups of alpha PVs based on the different E5 types they encode are shown. Inconsistencies in the distribution regarding different regions of the genome suggest a recombination event within the L2–L1 region for some of these PVs. The relationship between branch lengths and 0.1 matrix units is given by the bar at the bottom of each tree.

Box 1. Future directions and open questions

Which cellular proteins have coevolved with each of the different complexes L1–L2, E1–E2, E5–E6–E7?
Was there a coevolution of proteins from different viruses (other than PVs) targeting the same cellular proteins? If yes, how did they evolve?
Was there a common ancestor for high-risk and low-risk HPVs? If yes, how and when did the two lineages diverge, developing different mechanisms for the targeting of different proteins?
Can we learn from the evolution of the viral oncogenes which would be the best method to prevent uncontrolled growth?
Are there papillomaviruses:
in other clades within placental mammals?
in other clades within mammals (i.e. Marsupialia and Monotremata)?
in other clades within Amniota (i.e. turtles, lizards and crocodiles)?

non-human primates and in non-primates could prove to have both basic and clinical importance, considering the involvement of PVs in both malignant and benign transformations in humans (Box 1).

Hypothetical evolutionary scenario

There are five well-defined regions in the PV genome, with different evolutionary rates and histories. Assuming that PV classification must reflect biological relationships [41], our proposal of a composite nature of the PV genome brings into question the validity of the present narrow criteria for classifying PVs, strictly based on the sequence identity of a single gene [7].

We propose that the history of the PVs could have taken place in different steps. The first one corresponds to the initial appearance of the proto-PV, by combining pre-existing genetic information, leading to the URR-E1–E2-L2–L1 organization. This proto-PV would have infected an ancestor in the Amniota lineage. Further divergence processes broadened the spectrum of the proto-PVs. The proto-E6–E7 region later accessed to one of the proto-PVs, and evolved approximately twofold faster than the core region of the genome. Correlating with the emergence of the mammalian lineage 150 million years ago [47], there was a rapid diversification of the proto-PV repertoire, obvious today in the star-like appearance of the PV phylogenetic trees. As the interaction with the hosts became specific, a virus–host coevolution took place. PV infection mechanisms, which require direct contact between an infected and a non-infected host, fuelled this coevolution process. Different E5 ORFs most recently accessed the genome of the proto-alpha PVs and the proto-delta PV in fragile *loci* between E2 and L2. For the proto-alpha PV, this integration took place after the emergence of the ancestor of the primates and before the divergence within the primates family, ~24 million years ago. For the proto-delta PV, the integration of the E5 ORFs occurred after the apparition of the Artiodactyla ancestor ~65 million years ago and before the split Cervoidea/Bovoidea ~23 million years ago within this order of mammals.

Finally, for human PVs, phylogenetic analyses have provided the first hints connecting the biology of the viruses and differential association with malignant or benign transformations. This synergy between *in silico*, wet labour and epidemiological approaches highlights the importance of combined strategies for addressing complex problems.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science and Technology (BIO2003–07672), the BBVA Foundation (BIO04) and the Instituto Carlos III (C03/08).

References

1 zur Hausen, H. (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat. Rev. Cancer* 2, 342–350
2 Harwood, C.A. *et al*. (2004) Increased risk of skin cancer associated with the presence of epidermodysplasia verruciformis human papillomavirus types in normal skin. *Br. J. Dermatol.* 150, 949–957
3 Clifford, G.M. *et al*. (2003) Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br. J. Cancer* 88, 63–73
4 Kaya, H. *et al*. (2001) Prevalence of human papillomavirus (HPV) DNA in larynx and lung carcinomas. *Pathologica* 93, 531–534
5 Pfister, H. (2003) Human papillomavirus and skin cancer. *J. Natl. Cancer Inst. Monogr.* 31, 52–56
6 Doorbar, J. (2005) The papillomavirus life cycle. *J. Clin. Virol.* 32 (Suppl. 1), S7–S15
7 de Villiers, E.M. *et al*. (2004) Classification of papillomaviruses. *Virology* 324, 17–27
8 Chen, Z. *et al*. (2005) Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J. Virol.* 79, 7014–7023
9 Bravo, I.G. and Alonso, A. (2004) Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. *J. Virol.* 78, 13613–13626
10 Schiffman, M. *et al*. (2005) The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* 337, 76–84
11 Munoz, N. *et al*. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* 348, 518–527
12 Hughes, F.J. and Romanos, M.A. (1993) E1 protein of human papillomavirus is a DNA helicase/ATPase. *Nucleic Acids Res.* 21, 5817–5823
13 Dostatni, N. *et al*. (1991) The functional BPV-1 E2 trans-activating protein can act as a repressor by preventing formation of the initiation complex. *Genes Dev.* 5, 1657–1671
14 Demeret, C. *et al*. (1997) Different mechanisms contribute to the E2-mediated transcriptional repression of human papillomavirus type 18 viral oncogenes. *J. Virol.* 71, 9343–9349
15 Chen, X.S. *et al*. (2000) Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol. Cell* 5, 557–567
16 Holmgren, S.C. *et al*. (2005) The minor capsid protein L2 contributes to two steps in the human papillomavirus type 31 life cycle. *J. Virol.* 79, 3938–3948
17 Longworth, M.S. and Laimins, L.A. (2004) Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol. Mol. Biol. Rev.* 68, 362–372
18 Albá, M.M. and Castresana, J. (2005) Inverse relationship between evolutionary rate and age of mammalian genes. *Mol. Biol. Evol.* 22, 598–606
19 Fehrmann, F. and Laimins, L.A. (2003) Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene* 22, 5201–5207
20 Munger, K. *et al*. (2004) Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* 78, 11451–11460
21 Dyson, N. *et al*. (1989) The human papilloma virus-16 E7 oncoprotein is able to bind the retinoblastoma gene product. *Science* 243, 934–937
22 Caldeira, S. *et al*. (2003) The E6 and E7 proteins of the cutaneous human papillomavirus type 38 display transforming properties. *J. Virol.* 77, 2195–2206
23 Caldeira, S. *et al*. (2000) Human papillomavirus E7 proteins stimulate proliferation independently of their ability to associate with retinoblastoma protein. *Oncogene* 19, 821–826

24 Narechania, A. *et al*. (2004) Lack of the canonical pRB-binding domain in the E7 ORF of artiodactyl papillomaviruses is associated with the development of fibropapillomas. *J. Gen. Virol.* 85, 1243–1250
25 Huibregtse, J.M. *et al*. (1991) A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J.* 10, 4129–4135
26 Werness, B.A. *et al*. (1990) Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* 248, 76–79
27 Schiller, J.T. *et al*. (1984) Identification of a second transforming region in bovine papillomavirus DNA. *Proc. Natl. Acad. Sci. U. S. A.* 81, 7880–7884
28 Tachezy, R. *et al*. (2002) Avian papillomaviruses: the parrot Psittacus erithacus papillomavirus (PePV) genome has a unique organization of the early protein region and is phylogenetically related to the chaffinch papillomavirus. *BMC Microbiol.* 2, 19
29 Terai, M. *et al*. (2002) Lack of canonical E6 and E7 open reading frames in bird papillomaviruses: *Fringilla coelebs* papillomavirus and *Psittacus erithacus timneh* papillomavirus. *J. Virol.* 76, 10020–10023
30 O'Brien, V. *et al*. (2001) Cell transformation by the E5/E8 protein of bovine papillomavirus type 4. p27(Kip1), elevated through increased protein synthesis is sequestered by cyclin D1-CDK4 complexes. *J. Biol. Chem.* 276, 33861–33868
31 Bravo, I.G. *et al*. (2004) Human papillomavirus type 16 E5 protein. *Papillomavirus Rep* 15, 1–6
32 Bravo, I.G. *et al*. (2005) The E5 protein of human papillomavirus type 16 modulates composition and dynamics of membrane lipids in keratinocyte membranes. *Arch. Virol.* 150, 231–246
33 Chan, S.Y. *et al*. (1995) Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny and taxonomy. *J. Virol.* 69, 3074–3083
34 Delius, H. *et al*. (1994) Canine oral papillomavirus genomic sequence: a unique 1.5-kb intervening sequence between the E2 and L2 open reading frames. *Virology* 204, 447–452
35 Hughes, A.L. and Hughes, M.A.K. (2005) Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes. *Virus Res.* Doi: 10.1016/j.viruses.2005.03.030 (www.sciencedirect.com/science/journal/01681702)
36 Jackson, A. (2005) The effect of paralogous lineages on the application of reconciliation analysis by cophylogeny mapping. *Syst. Biol.* 54, 127–145
37 Milne, I. *et al*. (2004) TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* 20, 1806–1807
38 McGuire, G. and Wright, F. (2000) TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. *Bioinformatics* 16, 130–134
39 Bravo, I.G. and Müller, M. (2005) Codon usage in papillomavirus genes: practical and functional aspects. *Papillomavirus Rep* 16, 1–9
40 Zhao, K.N. *et al*. (2003) Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Res.* 98, 95–104
41 van Regenmortel, M.H.V. (2000) Introduction to the species concept in virus taxonomy. In *Seventh Report of the International Committee on Taxonomy of Viruses* (van Regenmortel, M.H.V. *et al*., eds), pp. 3–16, Academic Press
42 Van Ranst, M. *et al*. (1992) Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. *J. Gen. Virol.* 73, 2653–2660
43 Chan, S.Y. *et al*. (1992) Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. *J. Virol.* 66, 2057–2066
44 Bernard, H.U. (1994) Coevolution of papillomaviruses with human populations. *Trends Microbiol.* 2, 140–143
45 Bernard, H.U. (2005) The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *J. Clin. Virol.* 32 (Suppl. 1), S1–S6
46 Halpern, A.L. (2000) Comparison of papillomavirus and immunodeficiency virus evolutionary patterns in the context of a papillomavirus vaccine. *J. Clin. Virol.* 19, 43–56
47 Kumar, S. and Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. *Nature* 392, 917–920
48 Manski, K.M. *et al*. (1997) Bovine papillomavirus type 1 E1 and simian virus 40 large T antigen share regions of sequence similarity required for multiple infections. *J. Virol.* 71, 7600–7608
49 Garcia-Vallve, S. *et al*. (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10, 1719–1725
50 Narechania, A. *et al*. (2005) Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. *J. Gen. Virol.* 86, 1307–1313

Elsevier.com – Dynamic New Site Links Scientists to New Research & Thinking

Elsevier.com has had a makeover, inside and out. Designed for scientists' information needs, the new site, launched in January, is powered by the latest technology with customer-focused navigation and an intuitive architecture for an improved user experience and greater productivity.

Elsevier.com's easy-to-use navigational tools and structure connect scientists with vital information – all from one entry point. Users can perform rapid and precise searches with our advanced search functionality, using the FAST technology of Scirus.com, the free science search engine. For example, users can define their searches by any number of criteria to pinpoint information and resources. Search by a specific author or editor, book publication date, subject area – life sciences, health sciences, physical sciences and social sciences – or by product type. Elsevier's portfolio includes more than 1800 Elsevier journals, 2200 new books per year, and a range of innovative electronic products. In addition, tailored content for authors, editors and librarians provides up-to-the-minute news, updates on functionality and new products, e-alerts and services, as well as relevant events.

Elsevier is proud to be a partner with the scientific and medical community. Find out more about who we are in the About section: our mission and values and how we support the STM community worldwide through partnerships with libraries and other publishers, and grant awards from The Elsevier Foundation.

As a world-leading publisher of scientific, technical and health information, Elsevier is dedicated to linking researchers and professionals to the best thinking in their fields. We offer the widest and deepest coverage in a range of media types to enhance cross-pollination of information, breakthroughs in research and discovery, and the sharing and preservation of knowledge. Visit us at Elsevier.com.

Elsevier. Building Insights. Breaking Boundaries.